

Presented at the FIG e-Working Week 2021,  
21-25 June 2021 in Virtuality in the Netherlands



**e** WORKING WEEK 2021 20-25 JUNE

SMART SURVEYORS FOR LAND  
AND WATER MANAGEMENT

**CHALLENGES IN A NEW REALITY**

# Pattern Mining in Sentinel 2A Satellite Images Using Knime Analytics Platform

ID 10937

Rudiney Pereira, Elisiane Alba, Juliana Marchesan,  
Mateus Schuh and Roberta Fantinel



Federal University of Santa Maria  
Brazil



UFESM



**e**WORKING WEEK 2021 20-25 JUNE  
SMART SURVEYORS FOR LAND  
AND WATER MANAGEMENT  
**CHALLENGES IN A NEW REALITY**

# Introduction

- > The importance of research is to allow image processing without human interference to efficiently determine land use and land cover changes.
- > Land use patterns and land cover mining tool designed to manage knowledge from Sentinel 2A series satellite image data.
- > It was elaborated on Knime Analytics platform through the selection of configured and connected nodes constituting a workflow composing all the methodological phases.



**e**WORKING WEEK 2021 20-25 JUNE  
SMART SURVEYORS FOR LAND  
AND WATER MANAGEMENT  
**CHALLENGES IN A NEW REALITY**

# Introduction

> The graphic platform allows different configurations for the nodes with the main phases: loading and visualization of images; feature extraction (non-redundant numeric vector) that characterizes land use and cover; creation of attributes for each target (class of land use and land cover; test phase and evaluation of the predictive model.





**e**WORKING WEEK 2021 20-25 JUNE  
SMART SURVEYORS FOR LAND  
AND WATER MANAGEMENT  
**CHALLENGES IN A NEW REALITY**

# Introduction

- > The peculiarity in this case is that we must extract a vector of numerical resources from the image before starting the machine learning of the classifier algorithm in the implementation phase.
- > Two reasons justify:
  1. volume of images available
  2. is a variety of tools that are easy to use for data mining



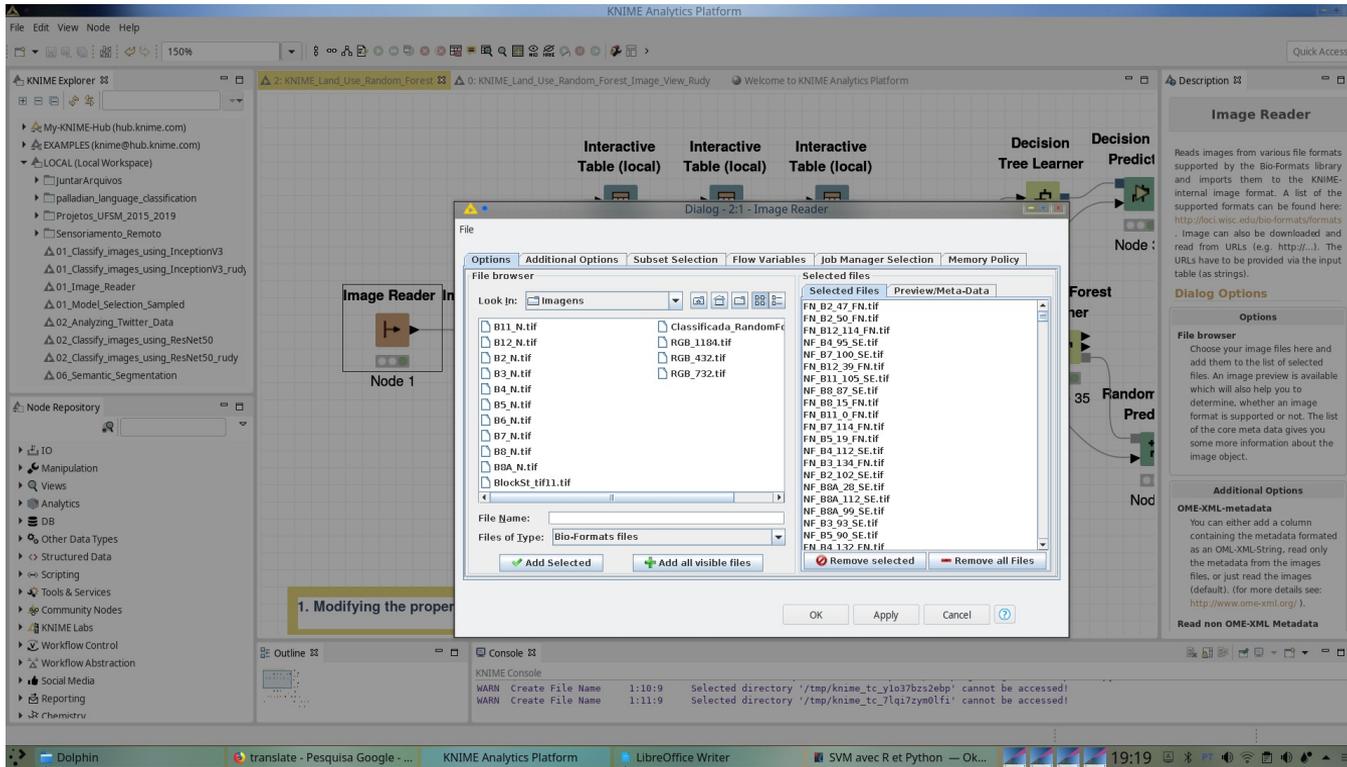


**e**WORKING WEEK 2021 20-25 JUNE  
SMART SURVEYORS FOR LAND  
AND WATER MANAGEMENT  
**CHALLENGES IN A NEW REALITY**

# Material and Methods

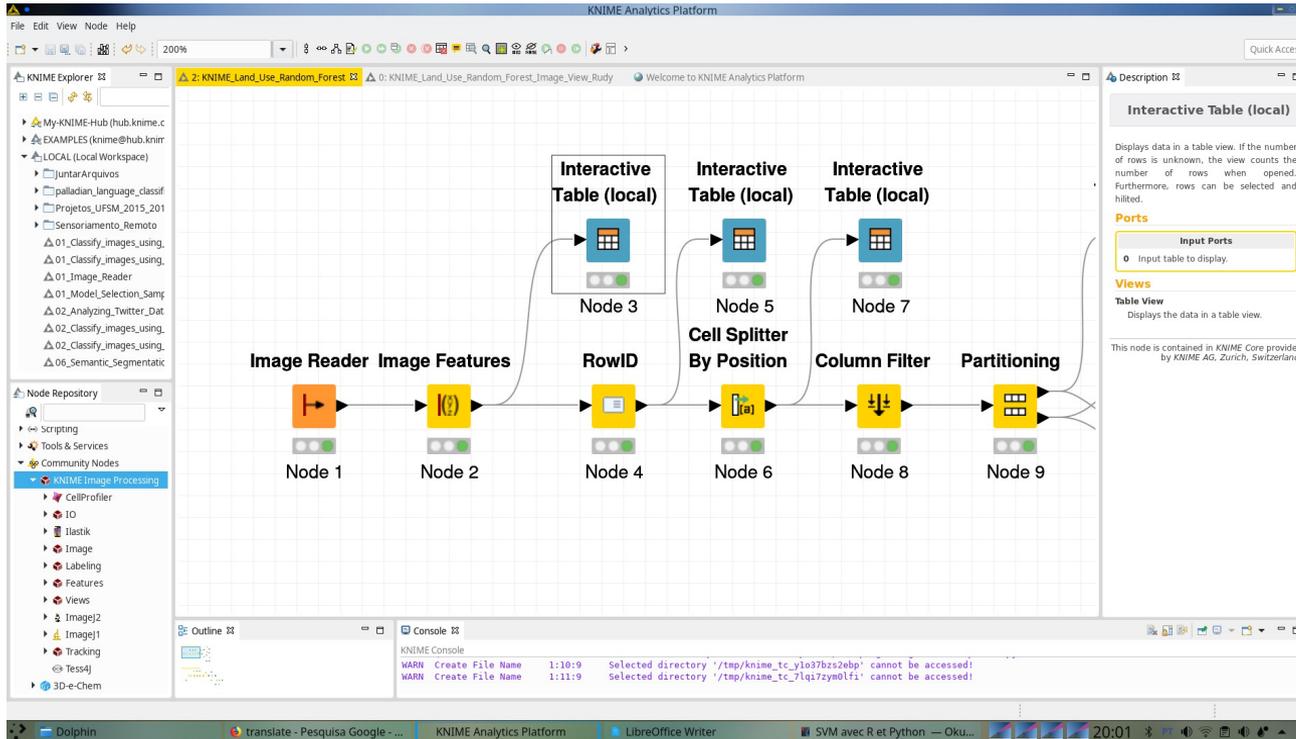
1. *Images dataset:* 1,370 files with 30 files per spectral band with 10 X 10 pixels of 10 spectral bands
2. *Building the Workflow in Knime:* was elaborated by placing nodes with their different functions and interconnected with each other to flow processing in main steps
  - . preparation of data such as reading, extracting characteristics and filtering data;
  - . data partitioning, machine learning in the decision tree and random forest algorithms and classification prediction;
  - . performance analysis of the classification algorithms.





Data sets:

1.370 images Tiff format  
 file



Building the workflow  
 in Knime Analytics  
 Platform, partial view



**e**WORKING WEEK 2021 20-25 JUNE  
SMART SURVEYORS FOR LAND  
AND WATER MANAGEMENT  
**CHALLENGES IN A NEW REALITY**

## Material and Methods

3. *Feature extraction*: was performed by the node called "Image Features" and consisted of configuring extraction options
4. *Creating the target attribute and learning and evaluating the predictive model*

**Machine Learning algorithm**

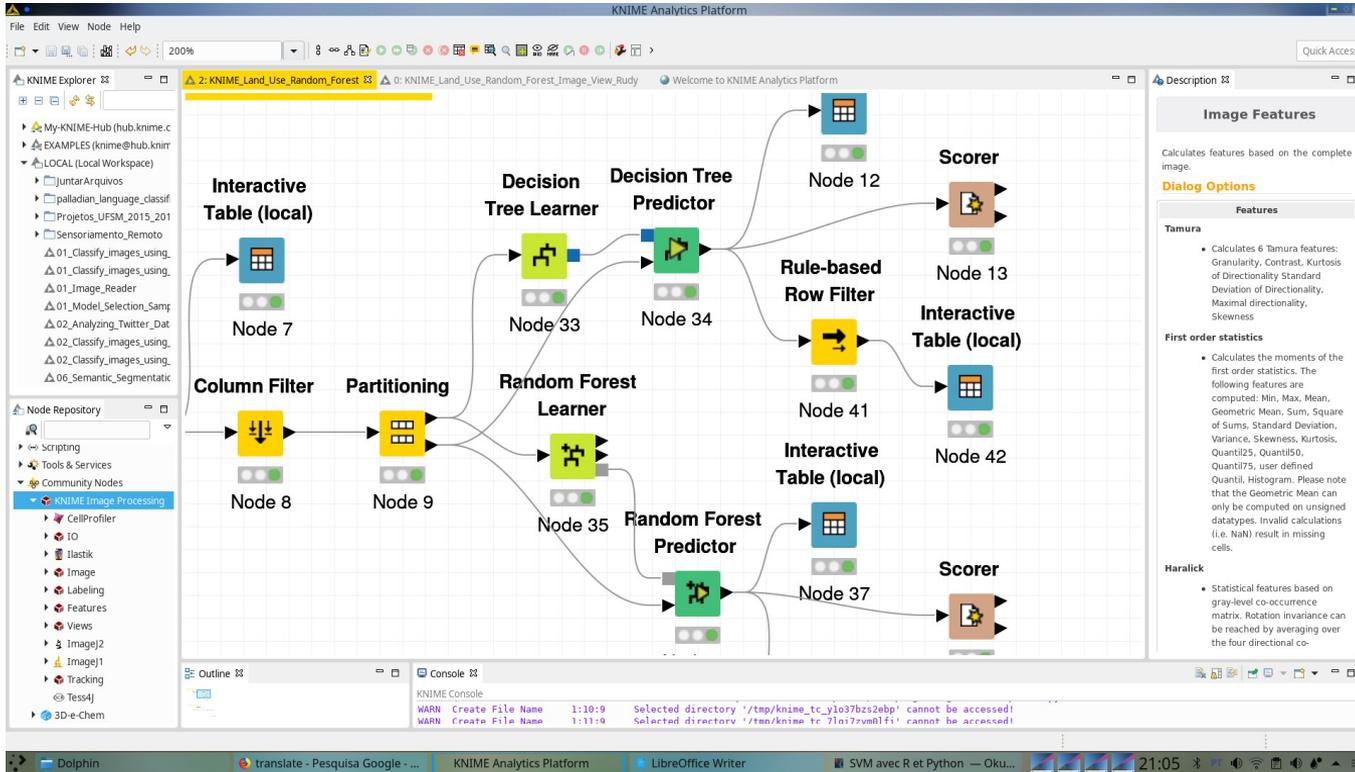
Interactive Table (local) Node 3  
 Interactive Table (local) Node 5  
 Interactive Table (local) Node 7  
 Cell Splitter By Position  
 Column Filter  
 Partitioning  
 Random Forest Node 33  
 Decision Tree Learner Node 34  
 Decision Tree Predictor Node 12  
 Rule-based Row Filter Node 13  
 Scorer Node 13  
 Interactive Table (local)

Row ID	ASM	Contr	Correl	Variance	IDFM1	SumAv	SumV	SumE	Entro	Differ	ICM1	ICM2	Cluste	Cluster	ASM	Contr
FN_B2_47...	0.007	146.148	0.243	96.955	0.104	36.049	246.072	3.812	5.011	0	3.145	-0.522	0.995	13.654.226.466.054.7...	0.007	77.877
FN_B2_90...	0.008	99.444	0.066	9.949	0.041	11.448	120.584	3.443	4.839	0	2.935	-0.438	0.966	19.764.226.650.653.4...	0.008	104.778
FN_B12_1...	0.009	79.802	0.444	71.788	0.221	37.407	207.353	3.668	4.818	0	2.792	-0.54	0.996	17.915.332.652.703.0...	0.008	53.395
Nf_B4_95...	0.01	57.679	0.45	52.442	0.168	26.247	152.087	3.363	4.698	0	2.671	-0.449	0.966	7.339.522.267.857.3...	0.011	24.296
Nf_B7_10...	0.008	87.457	0.475	83.358	0.161	36.543	245.977	3.589	4.908	0	2.893	-0.544	0.987	20.236.889.774.192.5...	0.008	36.284
FN_B12_3...	0.011	40.383	0.688	64.739	0.237	48.481	216.571	3.609	4.675	0	2.482	-0.557	0.986	48.116.915.2.165.897...	0.009	40.852
Nf_B11_1...	0.007	84.123	0.569	97.651	0.122	36.123	306.479	3.791	4.895	-0	2.784	-0.507	0.993	17.867.537.694.106.7...	0.01	40.457
Nf_B9_87...	0.008	60.988	0.5	60.955	0.114	30.79	182.832	3.584	4.919	0	2.579	-0.496	0.99	11.656.407.459.550.48	0.007	81.37
FN_B8_15...	0.008	100.494	0.396	83.128	0.162	50.173	232.02	3.629	4.891	-0	2.79	-0.536	0.986	46.756.076.2.040.444...	0.009	36.506
FN_B11_0...	0.008	67.741	0.661	99.835	0.107	36.21	331.588	3.746	4.869	0	2.684	-0.568	0.989	21.316.604.948.809.4...	0.009	61.296
FN_B7_11...	0.008	113.247	0.282	78.891	0.103	32.259	202.316	3.447	4.895	-0	2.895	-0.533	0.986	11.708.327.454.887.6...	0.007	78.827
FN_B5_19...	0.01	40.469	0.483	38.883	0.19	40.148	119.064	3.295	4.722	0	2.49	-0.423	0.959	26.613.194.1.042.517...	0.01	53.295
Nf_B4_11...	0.01	123.519	-0.029	59.996	0.235	28.827	116.464	3.454	4.732	-0	2.788	-0.432	0.963	5.836.692.168.702.5...	0.011	34.531
FN_B3_13...	0.01	42.16	0.505	42.623	0.137	30.037	128.332	3.404	4.757	-0	2.408	-0.43	0.962	10.300.948.381.167.5...	0.01	50.235
Nf_B2_10...	0.012	25.457	0.83	182.255	0.309	43.284	703.561	3.45	4.551	-0	2.015	-0.5	0.976	27.302.332.1.214.856...	0.011	100.839
Nf_B9A_2...	0.008	47.099	0.675	72.484	0.187	34.951	242.837	3.78	4.867	-0	2.484	-0.478	0.976	1.6.649.33.654.471.9...	0.009	38.689
Nf_B8A_1...	0.008	66.728	0.573	78.05	0.173	42.901	245.472	3.644	4.85	0	2.739	-0.534	0.995	29.863.243.1.191.173...	0.008	25.086
Nf_B8A_9...	0.01	56.531	0.533	60.557	0.199	45.074	185.698	3.567	4.792	-0	2.584	-0.446	0.968	35.293.525.1.436.740...	0.01	59.802
Nf_B3_93...	0.008	90.63	0.483	87.684	0.154	26.062	160.107	3.487	4.882	0	2.706	-0.521	0.984	8.086.282.315.381.6...	0.009	69.481
Nf_B5_90...	0.01	36.716	0.69	59.216	0.151	52.272	200.148	3.42	4.774	0	2.454	-0.507	0.98	60.704.803.2.686.402...	0.009	37.21
FN_B4_13...	0.009	112.348	0.113	63.309	0.084	40	140.889	3.351	4.786	0	2.892	-0.509	0.981	21.106.786.605.398.9...	0.009	111.049
FN_B4_19...	0.008	98.716	0.237	64.694	0.114	41.802	160.06	3.565	4.899	0	2.803	-0.512	0.983	24.675.832.950.452.8...	0.008	110.519
Nf_B5_78...	0.009	59.457	0.283	41.471	0.125	31.062	108.428	3.379	4.829	0	2.665	-0.434	0.965	9.911.01.312.130.2...	0.009	73.346
Nf_B8_89...	0.008	67.481	0.386	54.962	0.174	47.407	152.365	3.543	4.882	0	2.703	-0.476	0.976	41.823.237.1.710.244...	0.008	68.588
Nf_B8_10...	0.008	78.588	0.512	80.656	0.185	28.765	244.056	3.665	4.91	0	2.698	-0.504	0.982	7.926.305.287.147.06	0.009	50.358

## Feature extraction of images:

- . values minimum, maximum, average, geometric mean, standard deviation, variance, contrast, correlation, entropy and other parameters





Data set was partitioned into two other sets, training and testing

The data partitioning had two flows:

1. Decision Tree Predictor algorithm;
2. Random Forest Predictor algorithm

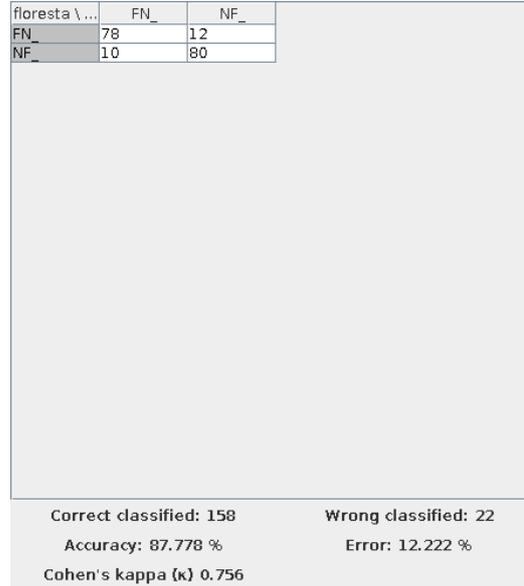
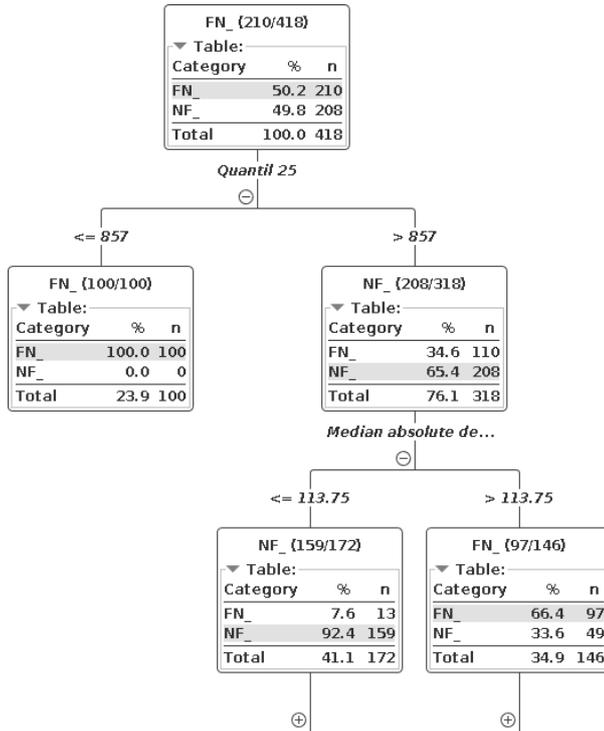
# Results and Discussion

**1. *Decision Tree algorithm:*** the node produces two information to analyze the classifier's performance.

*The first* consists of a graphical information in the form of a tree in which it presents the weights and percentages of each of the parameters, used in the characterization process of data image.

*The second* possible information is represented which shows the performance values of the classifier such as: samples classified correctly, classification errors and accuracy.

The Decision Tree classification algorithm had an accuracy of 87.778%, an error (incorrectly classified samples) of 12.22%, with the Cohen's Kappa  $k = 0.756$  index, considered satisfactory.



The **left figure** represent the classifier performance with weights and percentages of each of the parameters.

The **right figure** present Confusion Matrix.

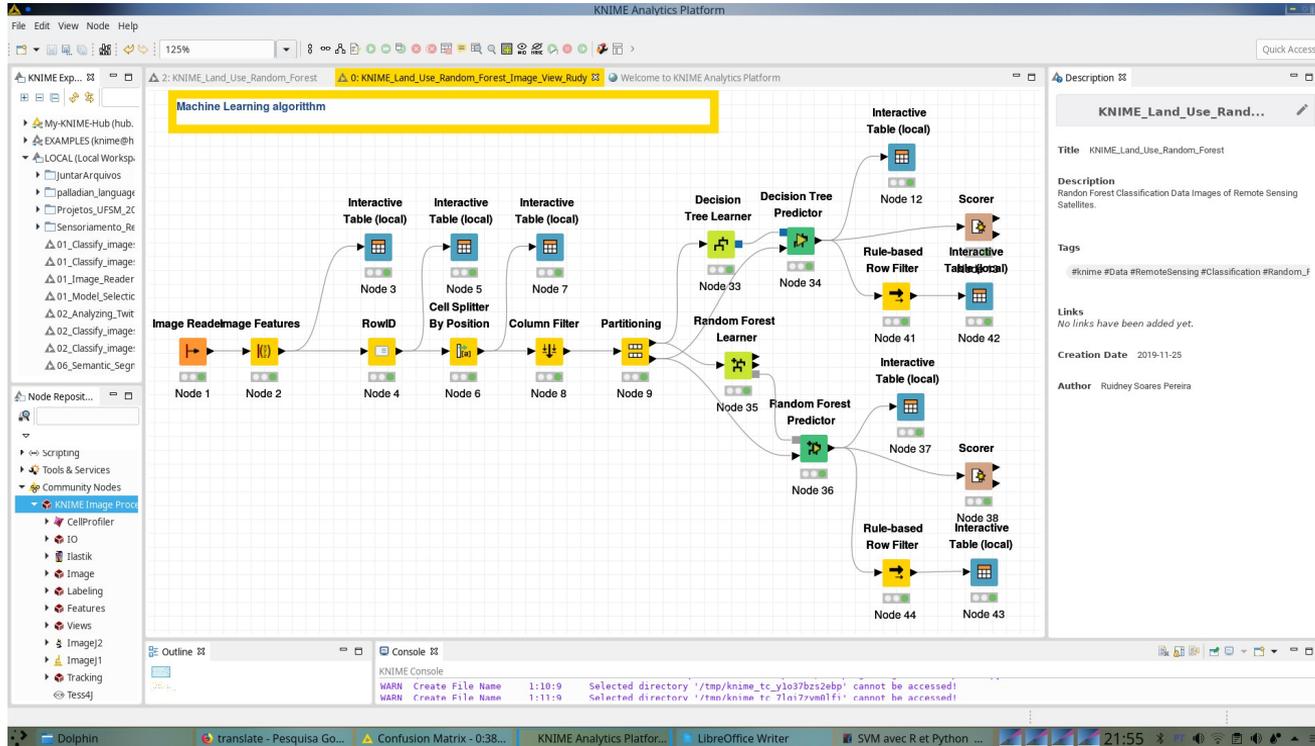


**e** WORKING WEEK 2021 20-25 JUNE  
SMART SURVEYORS FOR LAND  
AND WATER MANAGEMENT  
**CHALLENGES IN A NEW REALITY**

## Results and Discussion

**2. *Random Forest algorithm:*** shows the performance values of the classifier such as: samples classified correctly, classification errors and accuracy. The Random Forest classification algorithm had an accuracy of 93.333% accuracy, an error (incorrectly classified samples) of 6.667%, with the Cohen's Kappa  $k = 0.867$  index, considered satisfactory.





floresta \ ...	FN_	NF_
FN_	80	10
NF_	2	88

Correct classified: 168	Wrong classified: 12
Accuracy: 93.333 %	Error: 6.667 %
Cohen's kappa (κ) 0.857	

**Randon Forest Confusion Matrix Performance**

## Overview of Workflow Knime for image classification

# Conclusion

1. The Knime Analytics platform appears as a high-performance tool for complex analyzes without requiring a single line of code with a programming language;
2. The flow model used allows it to be improved, as it is possible to export, edit and adapt it to the interests of each user.
3. The analysis of the performance of the Decision Tree and Random Forest algorithms allowed us to conclude that it is possible to classify the images with the necessary precision.
4. Random Forest was the one that presented the best performance in the classification of images with the target of interest Forests.
5. Then, it is concluded that it is possible to make an intelligent knowledge management.



**e**WORKING WEEK 2021 20-25 JUNE  
SMART SURVEYORS FOR LAND  
AND WATER MANAGEMENT  
**CHALLENGES IN A NEW REALITY**

# Thanks for your attention!

Contact: [rudiney.s.pereira@ufsm.br](mailto:rudiney.s.pereira@ufsm.br)



Federal University of Santa Maria  
Brazil



UFSM